图书情报工作
第 64 卷 第 24 期 2020 年 12 月

国外知识组织体系管理工具比较分析与启示*

■ 纪姗姗 赵炳容 刘峥
中国科学院文献情报中心 北京 100190

摘 要：[目的/意义] 比较分析语义网技术发展环境下国外代表性知识组织体系管理工具的研究现状及关键功能，梳理相关研究的发展趋势。[方法/过程] 通过网络调研法，遴选国外 10 个代表性知识组织体系管理工具，从结构定义、项目维护、数据管理、互操作性、操作环境等方面比较分析其发展现状。在此基础上，为国内知识组织体系管理工具的建设工作提供建议。[结果/结论] 研究发现，国内知识组织体系管理工具的建设工作应依托语义网、关联数据、语义集成、语义互操作等技术促进功能创新，包括：灵活设计并配置多类型数据模型、支持可持续发展的管理与维护机制、支持用户分布式协同管理与控制、知识驱动的数据迭代式更新、面向语义的 KOS 重用与进化、建设以 KOS 为核心的知识组织框架等。

关键词：知识组织体系 知识组织体系管理工具 知识管理 词表 本体
分类号：G250
DOI：10.13266/j.issn.0252-3116.2020.24.009

1 引言

知识组织体系（Knowledge Organization Systems，KOS）又称知识组织系统，是用于知识组织的各类规范和方法的统称，是获取、利用知识的重要手段，不仅涵盖了传统的词典、术语表、分类表、主题词表和叙词表，还包括语义网络、本体等支持复杂语义描述的框架[1-2]。在数字信息管理与应用中，知识组织体系作为用户信息需求与知识资源之间的桥梁，可帮助用户快速发现所需的资源[3]。最初主要应用于图书情报领域，如支持文档标注及索引、查询扩展、术语提取、文本分类、翻译等服务，以提升对数字资源的访问[4]。而在语义网络环境下，尤其是从传统文件网络向具有结构化和富含语义的数据网络演进过程中，由于知识组织体系呈现出数据模型灵活、语义表达能力丰富等特点，它在各领域的知识组织管理、信息挖掘分析和开发利用海量信息资源实践中发挥日益重要作用，成为大型信息架构、内容管理、发现系统等应用的重要组件。

伴随着知识组织体系构建需求的增加，各类机构的知识管理人员亟需合适的工具支持知识组织体系的设计、维护与发布。知识组织体系管理工具（以下简称“KOS 管理工具”）可用于编制专业或通用的词表、本体，降低创建、开发、管理知识组织体系的时间和人力成本[5]。它在信息系统架构中起到一个中间件的作用，既服务于内容管理系统的元数据管理、知识分类、知识标引等模块，又可以应用于分类导航、语义检索、推荐系统、智能问答等信息服务系统（见图 1）。利用 KOS 管理工具，用户可以快速导入、转化、整合、修改已有的 KOS，还可以从零构建、编辑满足自身需求的 KOS，甚至还可以自动对文本语料进行分析并生成 KOS 的定制数据。

早在 20 世纪 90 年代，OCLC 等图情机构就推出网络数据库及视窗编辑管理系统，以支持 DDC、LCSH 等词表的编制[6]。随着知识组织体系类型及标准的发展，KOS 管理工具的功能也不断演化。除图情机构外，知识服务供应商等机构也推出多种类型工具以支持知识组织体系的管理[7]。在知识组织体系成为各类机构知识管理与知识服务研究重点的发展背景下，相关工具的分析与评价也成为十分有意义的研究内容。笔者重点对国外代表性 KOS 管理工具的关键功能进行评

* 本文系 NSTL“下一代国家科技创新知识服务开放系统”先期研发任务“STKOS 超级科技词表内容建设机制和发展研究（理学部分）”（项目编号：XQYF0101-1）研究成果之一，受“中国科学院文献情报中心成立七十周年主题论坛与纪念文集出版”项目资助出版。
作者简介：纪姗姗（ORCID：0000-0002-7834-5098），馆员，硕士；赵炳容（ORCID：0000-0002-5172-8785），馆员，硕士；刘峥（ORCID：0000-0002-2494-436X），副研究馆员，博士，通讯作者，E-mail：liuz@mail.las.ac.cn。
收稿日期：2020-10-05 修回日期：2020-12-24 本文起止页码：73-83 本文责任编辑：徐健

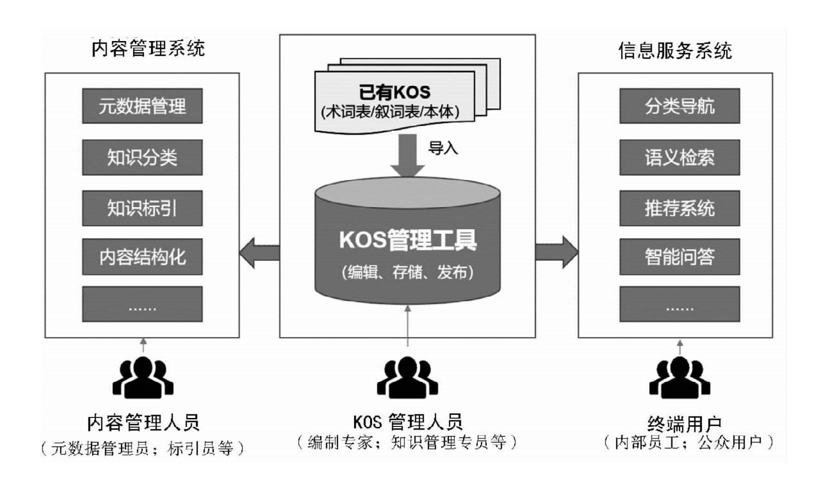


图 1 KOS 管理工具作用示意

估对比与现状分析，为国内各类机构及人员的知识组织体系建设工作提供建议。

2 相关研究

已有文献关于 KOS 管理工具的评估研究主要采用两种方法：

一是从工具建设者角度进行分析。如：J. Ganzmann[8] 从词表构建的通用需求及特定应用角度出发，列举了结构定义、数据输入、数据输出、检索能力等评估标准，并被广泛应用于叙词表构建时的工具选择；A. S. Chaudhry 等[9] 从准确性、灵活性、格式多样性、多语种支持、维护能力等方面对 Data Harmony 等 12 个词表构建工具进行评估；B. Pérez-León 等[10] 重点关注与语义网标准（特别是 SKOS）的兼容性以及集成或重用软件的能力，从支持的词表结构、数据格式、编辑功能、搜索功能、软件整合能力等角度对 Thmanager 等工具进行比较；P. Myrseth 等[11] 从模型标准、工作流、整合能力、易用性、信息治理、存储管理、发布、检索等多个角度出发，对 PoolParty 等 15 个词表及本体工具进行评价并打分；M. M. Martínez-González 等[12] 关注语义网环境下的词表管理工具的信息集成与语义互操作性，从用途、系统需求、功能、结构定义、一致性管理、信息互操作等角度对 MultiTes 等 6 个词表管理工具进行评估；白海燕[6] 提出知识组织工具管理软件的评价标准，包括结构定义能力、一致性控制能力、输出与共享交换能力、互操作支持能力、开放性与整合能力。最新的研究是 G. Mochón 等[13] 开展的开放关联数据技术发展下的词表管理工具评估，该研究制定了操作环境、概念及关系定义、创建与编辑、浏览与检索、互操作性等 5 个评估维度，细化为 27 条评估标准，对 TemaTres 等 10 个管理工具进行深度分析。

二是根据用户调研结果进行分析。如：A. Morshed 等[14] 设定了功能性、可重用性、数据存储、复杂性、可维护性、弹性、可靠性、鲁棒性、可学习性、可用性、灵活性和可见性等 16 项标准，依此收集 60 名用户的反馈数据，并基于无监督的机器学习算法进行计算，对 PoPoParty、Vocbench 和 TopBraid EVN 工具进行定量评估。此类方法相关研究较少，且评估标准偏主观性。

总结现有研究，KOS 管理工具的评估标准可概括为操作环境、结构定义、项目维护、数据管理、互操作性等 5 个维度：

（1）操作环境：较为通用的软件标准，用于分析软件的操作环境等因素，与知识组织体系的编辑无关，如：数据存储形式、应用类型、软件许可协议、兼容浏览器、操作系统、与第三方系统的交互与整合能力等。可参照标准如：ISO 25010、ISO 25012、ISO 25964-1 等。

（2）结构定义：工具遵循的数据模型标准，定义术语/概念/类的范围、属性及关系，决定知识组织体系的构建能力和复杂程度。细则可参照 ANSI /NISO Z39.19、ISO 25964-1、SKOS、OWL1/OWL2、ISO/IEC 21838 等标准。

（3）项目维护：对 KOS 的维护能力，如：是否支持多用户协同操作、活动报表、工作流管理等。

（4）数据管理：对 KOS 数据的编辑与管理能力，如：数据编辑（创建、删除、修改等）、数据浏览与检索、输入与输出、一致性与完整性控制等。

（5）互操作性：指知识组织体系之间的兼容、整合与互换能力，如：KOS 之间的数据映射、多语种 KOS 转换、KOS 数据整合等。

分析发现，现有的 KOS 管理工具评估研究多集中于对早期词表或本体建设工具的分析，并且尚未形成一个较全面的评估框架，而近年来知识组织体系建设

需求发展也促使 KOS 管理工具功能不断深化拓展。笔者将在现有研究的基础上制定较为全面的评估标准,对比分析 KOS 管理工具的基础功能及面向语义网发展的增值功能。

3　KOS 管理工具比较分析

3.1　KOS 管理工具调研与遴选

目前国内 KOS 管理工具的建设多基于特定任务开发,专业性、领域性较强,且使用环境相对封闭,不适用于通用性、开放性的知识组织体系建设,如 NSTL 科技知识组织体系素材管理系统[15]。国外 KOS 管理工具的研究较早,提供开放性的工具服务、有多领域用户应用案例,并且随着知识组织体系建设需求的发展不断动态更新。因此,此次选择国外有代表性的 KOS 管理工具作为调研对象。

为尽可能全面分析当前国外 KOS 管理工具的发展,笔者进行了广泛搜寻及严格筛选,范围涵盖 KMWorld 评选的知识管理领域 Top 100 公司产品[16]、相关文献研究中出现频率较高的产品、Taxonomy Boot Camp 推荐工具[17]等。最终遴选了 10 个 KOS 管理工具(见表1),它们具有以下特点:①功能以支持词表建设为主,同时也包括一些支持词表、本体等多类型知识组织体系建设的工具;②产品在近五年内发布,目前仍在提供服务且版本不断更新。

表 1　KOS 管理工具列表

产品	支持 KOS 类型	商业/开源	调研版本
Vocbench[18]	词表、本体	开源	Vocbench3
PoolParty[19]	词表、本体	商业	PoolParty Advanced Server 7.2
Synaptica KMS[20]	词表	商业	Synaptica KMS 9.0
Synaptica Graphite[21]	词表、本体	商业	Synaptica Graphite 3.0
TopBraid EDG-VM[22]	词表、本体	商业	TopBraid EDG 6.4
Mondeca ITM[23]	词表、本体	商业	2020
SKOS Shuttle[24]	词表、本体	商业	2017
TemaTres[25]	词表	开源	TemaTres 3.1
Unilexicon[26]	词表	开源	2019
MultiTes[27]	词表	商业	v2020.04.02u

注:调研版本说明:产品无版本标识的,以最新版本发布的时间为准

3.2　KOS 管理工具评估标准

笔者在现有 KOS 管理工具评估研究的基础上,结合近年来 KOS 管理工具功能的发展,从 5 个评估维度出发制定 18 条评估标准,对所调研的 KOS 管理工具进行全面深入的分析研究。具体评估标准如表 2 所示:

表 2　KOS 管理工具功能评估维度与标准

评估维度	评估标准	说明
结构定义	数据模型标准	遵循的数据模型标准,如:ANSI /NISO Z39.19、ISO 25964-1、SKOS、RDF/RDFS、OWL 等
	模型扩展	是否可以对基础模型进行语义扩展,如:词表模型中增加自定义概念属性与关系;建立"词表－本体"模型关联,利用本体扩展词表的语义类型,增强语义描述能力
项目管理	多语种管理	一个项目可选用两种以上语言对数据进行编辑
	状态管理	包括项目的编制状态以及项目中每条数据的加工状态
	版本管理	记录项目的版本变更情况,对比不同版本之间的内容变化
	历史记录管理	记录项目历史编辑信息,如最近一次的更新内容、时间以及责任人信息等数据
	用户权限管理	可设置不同用户或用户组对项目的访问权限
	工作流管理	管理 KOS 内容变化、控制其加工状态并管理发布周期
数据管理	编辑方式	除基础数据编辑功能外,工具可支持批量编辑、拖拽编辑等方式对数据进行编辑
	数据浏览	数据的呈现形式,包括字顺结构、层级结构、关系网络图等其他可视化方式
	数据检索	包括简单搜索、高级搜索、SPARQL 查询等方式
	数据质量控制	数据质量的审校与监管,实现内容的一致性与完整性控制
	数据更新	自动化从文本语料中挖掘相关的术语并补充至 KOS 内容中
互操作性	输入与输出	支持标准格式数据的导入和输出,实现与其它系统的数据共享,包括但不限于 CSV、RDF/XML、N-Triple、Turtle、N3、TriX、TriG、Zthes/XML 等格式数据
	KOS 内容映射	支持两个或多个 KOS 之间的内容映射
	开放关联数据集成	支持将 KOS 数据发布为开放关联数据,或者与 LOD 云中的数据建立关联关系
操作环境	数据存储形式	包括关系数据库、RDF 数据库等本地数据存储形式,以及外部数据库访问
	开放集成	与各种标引、资源管理系统或搜索引擎、检索工具整合,提供开放接口如 API 或通过 Web Services 供这些应用系统调用

3.3 对比评估与发展现状分析

基于以上评估标准,笔者通过产品文档调研、客服演示、工具试用、实例数据测试等方式对所遴选的 10 个 KOS 管理工具功能进行深入测试和对比评估,并根据各项评估结果分析当前国外 KOS 管理工具的发展趋势。

3.3.1 结构定义

(1)遵循多样化数据模型标准。对代表性 KOS 管理工具支持的数据模型标准进行分析,发现目前许多工具并不会着重区分“词表管理工具”与“本体管理工具”,更多的是既支持叙词表标准(ANSI/NISO Z39.19 或 ISO 25964-1)又支持本体标准(OWL 与 RDF),尤其是适用于各种受控词表的 SKOS 模型[7]。这种趋势是正常的,因为一个机构往往需要构建多种不同类型的知识组织体系。新兴的软件提供商已经将词表管理与本体管理功能结合在一起,包括 Vocbench、PoolParty、TopBraid EDG-VM、Mondeca ITM、SKOS Shuttle 等。Synaptica 是一个例外,它提供了两个产品:主要用于传统叙词表管理的 Synaptica KMS 以及主要用于本体管理的 Synaptica Graphite。具体如表 3 所示:

表 3 KOS 管理工具结构定义能力比较

评价标准		KOS 管理工具									
		Vocbench	PoolParty	Synaptica KMS	Synaptica Graphite	TopBraid EDG-VM	Mondeca ITM	SKOS Shuttle	TemaTres	Unilexicon	MultiTes
数据模型标准	Z39.19										√
	ISO 25964-1			√							
	RDF/RDFS	√	√		√	√	√	√	√	√	
	SKOS	√	√		√	√	√	√	√	√	
	OWL	√	√		√	√	√	√			
模型扩展	自定义属性/关系	√	√	√	√	√	√	√			√
	基于本体的词表模型扩展	√	√		√	√	√	√			

(2)灵活性、可定制的模型扩展方式。大数据环境下,随着知识组织体量的扩大,实体之间的关系表达也日益丰富。例如,一般的叙词表中的关系仅包含用、代、属、分、参、族等,属性包括定义、范围注释、历史注释、一般注释等,一般都不超过 10 种。但是,新型知识组织体系如 UMLS 的语义网络中就包含了 54 种语义关系[28]。基础的叙词表数据模型已不足以支撑新型知识组织体系的语义表达,需要进行扩展。常见方式是在词表模型的基础上增加自定义数据属性或关系属性(参照 ISO 25964-1 标准)。除 TemaTes、Unilexicon 外,目前主流的 KOS 管理工具都支持对词表概念进行自定义属性/关系扩展。

此外,许多 KOS 管理工具开始支持建立“词表-本体”模型之间的关联,利用本体模型扩展词表模型的语义类型、属性及关系[29]。此种方式得益于 SKOS 与 OWL 的兼容性。例如,PoolParty 提供自定义知识模型的功能,建立词表内容与本体的关联,在语义层实现对词表模型的扩展描述,为单一的“概念”类型增加新的本体类,并继承该本体类所有的属性及关系,实现流程如图 2 所示。TopBraid EDG 也提供类似功能,它支持在本体管理中对词表模型(SKOS、SKOS-XL)进行扩展描述,为概念类(skos:concept)添加新的子类(如 country、city 等),为其添加属性及关系,并将扩展后的模型复用于原词表。

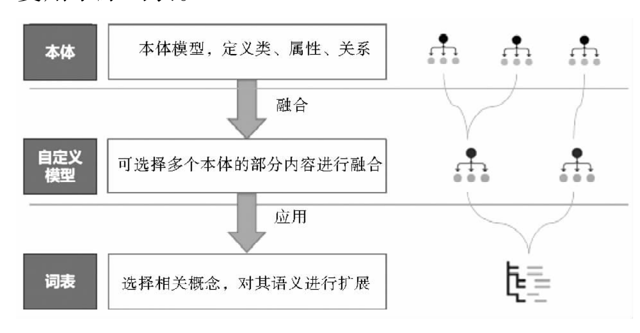


图 2 PoolParty 基于本体的词表模型扩展方法示意

3.3.2 项目维护

知识组织体系的构建过程中往往需要同时管理多个项目,以及基于网络环境的多用户协同编辑。除基础的项目浏览、项目注册功能之外,KOS 管理工具也在多语种管理、状态管理、版本管理、历史记录管理、权限管理、工作流管理等方面不断提升细节化的处理。

(1)支持国际化发展的多语种管理。随着不同文化和语言背景的信息交流需求的增加,支持多语种信息组织和检索的工具也日益重要,多语种是知识组织体系发展的一个重要趋势,这也促使 KOS 管理工具提供相应的管理功能。除 TemaTres、Unilexicon 在一个项目中只能选择一种语言对数据进行编辑之外,大部分 KOS 管理工具都支持多语种知识组织体系的数据编

表 4　KOS 管理工具项目维护能力比较

评价标准		KOS 管理工具									
		Vocbench	PoolParty	Synaptica KMS	Synaptica Graphite	TopBraid EDG-VM	Mondeca ITM	SKOS Shuttle	TemaTres	Unilexicon	MultiTes
多语种管理	多语种编辑	√	√	√	√	√	√	√			√
	自动翻译		√								
状态管理	数据加工状态	√	√	√	√	√	√	√			√
历史记录管理	变更信息记录	√	√	√	√	√	√	√			√
	历史记录检索	√	√			√					
版本管理	版本关联	√	√	√		√					
	版本对比			√		√					
用户权限管理	用户角色权限	√	√	√	√	√	√	√		√	
	项目权限控制	√	√			√					
工作流管理	审批工作流	√	√	√	√	√	√	√			

辑，如 Vocbench、PoolParty 等工具支持用户设置 KOS 数据浏览与编辑时的主语言以及其他偏好的语种，并为其排列显示次序。此外，PoolParty 还支持基于 DBpedia 实现对优选词的自动翻译，翻译结果可选择添加到相关的优选词或非优选词字段。

（2）维护项目可持续发展的状态、历史记录与版本管理。知识组织体系的状态管理需要分层次进行，对整个 KOS 和 KOS 中的单个术语应采用不同的状态管理策略。在项目层面，知识组织体系的状态指 KOS 整体内容所处的生命周期阶段，可采用开发中、维护中、已废弃等状态信息进行描述；在数据层面，状态管理主要对 KOS 中的单个术语的状态（如活动中、已废除等）进行描述。目前多数 KOS 管理工具仅支持在数据层面进行状态管理，不支持项目层面的状态管理。

NISO 建议，词表管理系统应当自动记录数据最近一次的更新内容、时间以及责任人信息，以保证任何变更信息都能够进行追溯和验证[30]。除 TemaTres、Unilexicon 之外，多数 KOS 管理工具提供历史记录管理功能，自动记录每条数据最近一次的更新内容、时间以及编辑人员等数据，保证 KOS 项目的任何变更信息都能够进行追溯和验证。Vocbench、PoolParty、TopBraid EDG-VM 还支持对历史记录进行检索和筛选，便于直观发现特定时间段、特定术语的信息变更。

此外，知识组织体系的建立不是一劳永逸的，需要有动态更新机制，形成对多个版本的分类与管理。版本控制的概念在知识组织体系维护中非常重要，制定良好的版本控制策略使得内容更新更容易，并且更适用于自动化的流程。目前 Vocbench、PoolParty、Synaptica 及 TopBraid EDG-VM 允许授权用户生成新版本的 KOS 数据，并建立不同版本之间的关联。TopBraid EDG-VM 还支持对不同版本的 KOS 内容进行对比，形成数据对比报告。

（3）支持用户协同工作的权限与工作流管理。知识组织体系的建设过程中可能涉及项目管理员、各领域词表专家、最终用户的大规模协同工作，因此需要为不同用户授予不同的访问权限，避免因权限控制缺失或操作不当引发的风险问题。最简单的方式是为不同用户或工作组分配不同的角色，通过角色控制用户对 KOS 数据的访问、编辑、审核、管理等权限，这种方式也被多数 KOS 管理工具采用。另一种方式是在角色控制的基础上，为用户设置每个项目的访问权限，这种方式更为严谨。例如：Vocbench 除基础的用户角色管理之外，每个项目都包含一个访问控制列表，包含 < ProjectConsumer, AccessLevel > 实体对，AccessLevel 可以设置 R（读）或者 RW（读/写）两个值，对用户的访问权限进行详细配置。

为便于管理 KOS 内容变化、控制其加工状态并管理发布周期，往往需要引入认证工作流，对用户的各项操作进行核查，这一机制与 KOS 项目的状态管理、版本管理、历史记录管理、权限管理等紧密结合。目前许多 KOS 管理工具有工作流管理功能，如：Vocbench 在其数据层构建了一个追踪改变装置，当数据发生变化（删除、添加）时，会同时保存操作元数据，授权用户可以选择接受或者拒绝某个操作；PoolParty 提供了一个简单的审批工作流，概念被创建、编辑、导入或通过推荐系统添加时，其默认状态为“草稿”，经审查人员批准后，状态会自动修改为“批准”；TopBraid EDG-VM 通过认证工作流对数据变更进行控制，用户对 KOS 项目内容进行编辑时，会先生成一个副本，用户所有的编辑操作都在此副本上进行，只有当副本中内容的变化被审查

人员通过后,相关变化才会同步到原 KOS 数据中。

3.3.3　数据管理

数据管理是 KOS 管理工具中最核心的功能模块,可以实现 KOS 数据的浏览、编辑、检索、质量控制、更新等操作。具体管理功能模块比较如表 5 所示:

表 5　KOS 管理工具数据管理能力比较

评价标准		KOS 管理工具									
		Vocbench	PoolParty	Synaptica KMS	Synaptica Graphite	TopBraid EDG-VM	Mondeca ITM	SKOS Shuttle	TemaTres	Unilexicon	MultiTes
编辑方式	批量编辑		√	√	√	√	√	√	√		
	拖拽编辑		√	√	√	√					
	自动拼写建议		√								
数据浏览	层级结构	√	√	√	√	√	√	√	√	√	√
	关系网络图	√	√		√	√	√	√		√	
	其他形式	√	√	√	√					√	
数据检索	简单检索	√	√	√	√	√	√	√	√	√	√
	高级检索	√	√	√	√	√	√	√	√		√
	SPARQL 查询	√	√		√	√	√	√			
数据质量控制	检测报告	√	√	√	√	√	√				√
	实时检测	√	√	√	√	√	√				
	自动修复	√									√
数据更新	基于文本语料的术语抽取		√					√			

(1)多样化数据编辑、浏览与检索形式。为提升数据编辑工作效率,多数 KOS 管理工具支持批量编辑管理功能。同时,PoolParty、Synaptica、TopBraid 公司产品支持通过拖拽形式实现数据的层级结构变化。PoolParty 还可以基于 DBpedia 数据在用户输入时提供自动拼写建议。

KOS 数据的可视化呈现,可以让终端用户能对其结构和内容有一个丰富的、形象化的理解,随时审视编制中的数据并反馈修改建议。不同工具也衍生出多种可视化形式,如层级结构、关系网络、可视映射图等。①层级结构:又可称为树形结构,这是 KOS 管理工具提供的最基础的数据可视化形式,按照 KOS 数据中的“上位概念 – 下位概念”或“父类 – 子类”关系显示所有数据,几乎所有 KOS 管理工具都支持层级结构的数据展示方式。②关系网络图:多数 KOS 管理工具支持通过“节点 – 关系链接”来显示数据关系,如 SKOS Shuttle、Mondeca ITM 等。③其他形式:如 PoolParty 采用可视化映射图显示概念的所有的关系,并用不同的颜色进行区分;Synaptica KMS 通过饼图显示数据分布情况;Synaptica Graphite 可通过项目的结构视图显示不同 KOS 模型之间的关联关系;Unilexicon 通过卡片形式支持用户对数据层级进行调整、排序。

同时,多数 KOS 管理工具支持简单检索与高级检索功能。高级检索时可对标签、关系、属性等内容进行检索,支持多种匹配模式(如精确匹配、模糊匹配、起始字符、结束字符、包含字符等),并且与数据筛选功能进行结合。Vocbench、PoolParty 等工具还支持 SPARQL 语言查询功能,便于有经验的用户快速定制检索语句。

(2)嵌入管理流程的数据质量控制。知识组织体系建设工作离不开对数据内容及其质量的审校与监管,这一工作往往需要人工的参与,但优秀的 KOS 管理工具可以提供多种数据质量检测方案,辅助人工审校过程,提升知识组织体系的构建效率。总体来讲,KOS 数据质量检测主要分为结构检测与标签检测两个方面:结构检测用于验证数据结构的完整性,如关系缺失或逻辑循环错误等;标签检测重点关注知识对象描述的完整性与一致性,如词形重复性、信息缺失等。KOS 管理工具对质量管理的支持存在两个层面:①在 KOS 数据导入后或完成一个阶段的加工任务后,对 KOS 数据进行整体的质量问题检测,并提供检测报告;②在编辑 KOS 数据时实时进行检测,将发现的错误实时反馈给用户。如:PoolParty 通过集成 qSKOS 来支持数据的质量管理,它可以设置许多质量评估标准,并在用户交互过程中实时检测;TopBraid EDG-VM 通过 SHACL(Shapes Constraints Language,结构性约束语言)[31]实现完整性问题的实时检查。有些工具如 Vocbench 在检测到错误之后还可以提供操作建议,用户可一键式实现标签或结构错误的自动修复。

(3)基于机器学习的数据更新。随着自然语言处理、机器学习技术的发展,实现 KOS 内容的智能化更

新成为当前研究热点和发展趋势，一些 KOS 管理工具也在研究如何从文本语料中挖掘相关的术语并补充至 KOS 内容中。例如：①PoolParty 支持从文本语料中自动挖掘相关的术语/概念，并将其作为候选词汇添加到词表中。术语抽取过程中采用互信息得分(Mutual Information Score, MIS)、内容术语分数（Content Term Score, CTS）、术语频率等筛选指标进行综合计算，得到候选术语[32]。②SKOS Shuttle 支持基于多语种(德语、英语、西班牙语、法语)语料的信息抽取，并使用 tf-idf 值对每个术语进行加权。用户可对抽取的结果进行筛选、排序，同时可选择将识别到的新术语添加到已有词表[33]。

3.3.4 互操作性

知识组织体系的互操作是当前 KOS 管理工具发展的重要特征，可实现从单一知识组织体系的独立建设向各类知识组织体系的整合发展，从单一领域的局部建设向跨领域的集成发展。相应的实现方法包括输入与输出、KOS 内容映射、开放关联数据集成等。标准格式数据的输入和输出是一种最基础的互操作方式，也是衡量 KOS 管理工具共享和交换能力的重要因素。常见的格式如：Txt、Excel/CSV、SKOS/SKOS-XL、RDF/XML、N-Triples、Turtle、JSON 等。各工具对不同格式数据的支持性如表 6 所示：

表 6　KOS 管理工具互操作性比较

评价标准		KOS 管理工具									
		Vocbench	PoolParty	Synaptica KMS	Synaptica Graphite	TopBraid EDG-VM	Mondeca ITM	SKOS Shuttle	TemaTres	Unilexicon	MultiTes
输入与输出	Txt			→ ←					←	→	→ ←
	Excel/CSV		→	→ ←		→			→	←	←
	SKOS/SKOS-XL		→ ←		→ ←	→ ←		→ ←	→ ←	←	←
	RDF/XML	→ ←	→ ←	→ ←	→ ←	→ ←	→ ←	→ ←			→ ←
	N-Triples	→	←		→ ←	→ ←	→ ←	→ ←			
	Turtle	→	←		→ ←	→ ←	→ ←	→ ←			
	JSON/JSON-LD	→	←		→ ←	→ ←	→ ←	→ ←		←	
KOS 内容映射		√	√	√	√	√		√			
开放关联数据集成		√	√	√	√	√		√			

说明："→"代表数据输入，"←"代表数据导出

(1)半自动化 KOS 内容映射。随着数据库集成应用与发展，出现了跨库、跨语言、跨领域的操作需求，不同知识组织体系之间的互操作成为急需解决的重要问题。目前解决互操作的最有效方法是实现 KOS 间的映射，即在两个或多个 KOS 间建立词汇对应关系，使其所标引的结果间也能进行关联。多数 KOS 管理工具支持通过机器自动识别的方式进行半自动化映射，例如：Vocbench 支持通过基于标签的半自动化搜索建立不同 KOS 项目之间的映射，相关实体与关系经验证之后采用标准的 RDFS/OWL 或 SKOS 映射属性建立关联，如两个 SKOS 概念可以通过 SKOS：exactMatch 或 SKOS：closeMatch 对齐。

(2)面向开放关联数据的集成。一方面支持将词表发布为开放关联数据，另一方面支持从 LOD 数据集中获取数据，扩展当前知识组织体系内容。例如 PoolParty、TopBraid、Vocbench 等工具可以将平台中的 KOS 内容在网络上发布为关联数据，并且与其他开放数据集(如 DBPedia、WordNet、LCSH、GeoNames)中的资源建立关联关系，从而成为关联开放数据(LOD)云的一部分。Vocbench、PoolParty 还支持自动与 DBPedia 等 LOD 数据进行匹配，将发现的相关数据自动添加至 KOS 内容中。

3.3.5 操作环境

KOS 管理工具操作环境比较见表 7。

(1)支持分布式语义数据存储。随着知识模型的演变以及数据量的激增，KOS 管理工具应当支持大规模、分布式语义数据的存储与管理。早期的 KOS 管理工具多数以关系数据库存储数据，如 TemaTres、Unilexion、

表 7 KOS 管理工具操作环境比较

评价标准		KOS 管理工具									
		Vocbench	PoolParty	Synaptica KMS	Synaptica Graphite	TopBraid EDG-VM	Mondeca ITM	SKOS Shuttle	TemaTres	Unilexicon	MultiTes
数据存储	关系数据库			√					√	√	√
	RDF 数据库	√	√		√	√	√	√			
	远程数据库访问	√	√			√	√	√			
开放集成	第三方 API 调用/Web 服务访问	√	√	√	√	√	√	√		√	
	与配套工具的集成		√			√	√				

MultiTes 等。现在，RDF4J、AllegroGraph、Jena TDB、Virtuoso、GraphDB、StarDog、MarkLogic 等 RDF 数据库逐渐成为主流 KOS 管理工具的首选，例如 Vocbench、PoolParty 选择 RDF4J 作为默认的本地语义仓储。除本地数据存储之外，一些 KOS 管理工具还支持远程访问其他存储，实现数据的分布式管理，如：Vocbench、PoolParty 等支持访问、调用 MarkLogic、Stardog、Allegrograph、GraphDB、Virtuoso 等远程存储的数据。

（2）与其他应用的无缝集成。目前，越来越多的 KOS 管理工具开始支持与内容管理系统及搜索软件相结合，将 KOS 数据嵌入到内容标注、信息检索、知识发现系统平台中。例如：Synaptica 产品可以与多种第三方应用相结合，包括 SharePoint、Google Search、Apache Solr、OpenText、MarkLogic 以及 IBM Watson 等，扩大应用范围。而 PoolParty、TopBraid、Mondeca 还提供相关配套组件，支持用户按需集成。如：PoolParty 语义套件包括词表与本体管理、知识图谱管理、文本挖掘/自然语言处理、关联数据编制、语义智能应用等功能组件，可根据用户需求自由组合集成；Mondeca ITM 可以与内容标引管理工具集成，并从中提取实例对 KOS 数据进行扩充。这一发展极大提升 KOS 管理工具的开放性与整合能力，实现与其他应用或管理系统之间的动态、实时的交互，进而扩大和提升 KOS 的应用范围和实用价值。

4 启示

研究发现，为支撑大数据环境与智能应用，国内 KOS 管理工具的建设不仅需要支持基础的 KOS 生命周期管理（创建、评估、发布、维护、进化），还应借鉴先进 KOS 管理工具的发展现状，结合实际知识组织体系构建需求，重点从以下几个方面提升工具性能：

4.1 灵活设计并配置多类型数据模型

知识组织体系的构建要考虑的是“如何合理有效地表示知识”，即创建一个标准化、抽象化的模型来描述现实世界中的对象，支持机器理解。对于叙词表、分类表、术语表等 KOS，其模型中基础的元素为术语或概念；而对于本体或知识图谱来讲，其知识模型中包含了类、类的属性、类之间的关系及相关的规则[34]。随着知识图谱研究热潮的兴起，如何对传统知识组织体系进行语义扩展，快速构建轻量级本体，成为 KOS 管理工具亟需解决的问题。工具应当提供不同的知识组织方案来支持灵活的数据建模，包括：基于术语的词表、基于概念的词表、描述复杂语义关系的本体等。具体方案设计可以参照 SKOS/SKOS-XL、ISO-THES[35]、RDF/XML 等模型标准进行扩展。

4.2 支持可持续发展的管理与维护机制

NISO 指出，随着时间推移，知识组织体系的访问需要足够的文档、合适的维护策略、积极的管理[30]。为实现知识组织体系的可持续性发展，需要对各种类型知识组织体系提供权威的、集中化的存储与管理，支持 KOS 的发现与重用，包括状态管理、权限管理、历史记录与版本管理等。所调研的 KOS 管理工具中，多数重点关注微观层面的数据状态管理，并不支持宏观层面的项目状态管理。因此，KOS 管理工具应当对项目的开发、维护、发布等状态提供更加细化的管理方案。

4.3 支持用户分布式协同管理与控制

当前知识组织体系的建设具有数据规模大、覆盖领域广泛等特点，这也决定其构建过程需要多领域多机构多层次用户协同编制和维护[36]。这就需要 KOS 管理工具：①支持分布式协同工作，便于不同群体用户协作完成 KOS 设计与编辑工作，提高工作效率；②具备灵活的权限和任务管理机制，为不同权限的用户分配不同层级的任务（如编辑、审核等），并支持进度管理。所调研的 KOS 管理工具中，虽大多数支持不同权限用户的协同化编制，但并不具备完善的从任务分配、数据编辑、数据审核到项目发布的全流程任务管理功能。KOS 管理工具建设时，应根据实际的项目协同工作需求，提供不同层级的角色权限，支持自定义工作

流，提供完整的任务管理功能。

4.4 实现知识驱动的数据迭代式更新

知识组织体系的构建是不断迭代的，在大数据时代，如何尽可能地节省人力物力，采用机器学习等先进技术对 KOS 数据进行智能化更新已成为重要的研究课题[28]。目前 PoolParty、SKOS Shuttle 已支持基于文本语料的术语抽取，但算法相对简单，仅通过词频、互信息等统计算法判断术语的相关性，且并未实现关系的挖掘。为实现知识组织体系的快速构建，KOS 管理工具应当拓展在数据挖掘方面的性能，支持搭载机器学习、深度神经网络等多种算法模型实现术语及关系的抽取。

4.5 面向语义的 KOS 重用与进化

语义网鼓励数据共享和数据重用，对已有术语的重用能够最大化 KOS 的使用价值，扩大其传播范围并保持生命力。例如：通过翻译、映射、合并、定制等方式，充分利用 KOS 管理工具中已有的数据构建新的知识组织体系，支持根据用户需求动态集成与分解。目前 KOS 管理工具提供内容映射以及开放关联数据的集成功能，但对于现有数据的动态化集成与定制支持性较弱。未来 KOS 管理工具可以通过以下方式拓展数据重用方式：①提取不同 KOS 中的类或概念组进行数据整合，涉及同义术语合并、关系融合等操作；②上传领域关键词并同时与多个 KOS 数据进行匹配，筛选匹配数据进行整合，涉及术语权重计算、关系扩展等操作。以上方式可支持跨领域知识组织体系的快速构建。

4.6 建设以 KOS 为核心的知识组织框架

KOS 管理工具的发展应当注重互操作、开放性和集成能力，支持与内容管理系统、标注及检索系统相整合，形成一体化的工作流程，便捷的应用于各类型机构的知识管理业务。机构知识组织建设工作中，应当将 KOS 管理工具作为核心组件，支持各类型数据资源的结构化与分类、知识实体的抽取与索引、检索语句的理解与分析等功能，促进 KOS 数据的应用，最大化发挥其价值。

此外，从 KOS 管理工具使用者的角度来看，不同 KOS 管理工具的功能存在较大差异。开源软件中：Vocbench 各方面表现均十分突出，满足语义网环境下的各类型知识组织体系构建，但它与第三方系统的集成主要通过编程方式调用，且操作较为复杂，适合具有词表、本体等 KOS 编制经验的用户；TemaTres、Unilexicon 功能较为单一，仅支持词表类知识组织体系建设，能满足基础的项目管理及数据管理，但不支持语义扩展、语义互操作等性能，适合小规模词表建设需求的用户。商业软件相对功能较为完善且有良好的服务保障，用户可根据自身数据建设规模与应用需求，选择合适的 KOS 管理工具。如：①普通词表编制用户可选择 MultiTes 或 Synaptica KMS 工具，其中 Synaptica KMS 具备良好的用户权限与工作流管理功能，并且支持拖拽编辑与数据内容的图形化展示，实现 KOS 数据的快速配置。②具有多类型知识组织体系构建需求的用户可选择 Synaptica Graphite、Mondeca ITM 或 SKOS Shuttle。它们功能相似且各有特色：Synaptica Graphite 支持拖拽编辑，以及数据或项目不同层面的可视化展示方式；Mondeca ITM 可以与其内容标引管理工具集成，并从中提取实例对 KOS 数据进行扩充；SKOS Shuttle 支持基于文本语料的术语抽取，实现 KOS 的智能化更新。③企业级用户可选择 PoolParty 或 TopBraid EDG-VM，它们拥有大量的客户应用案例，可处理各种类型复杂化、数据规模化的词表和本体，同时提供配套的内容标引、自动分类、图谱数据管理、知识发现等组件，实现与 KOS 管理工具的无缝集成，适合企业级知识组织体系构建与管理。

5 结语

KOS 管理工具的建设是实现机构数据资源有序组织、集成与服务的核心。为了解 KOS 管理工具的发展现状，笔者从结构定义、项目维护、数据管理、互操作性、操作环境 5 个维度出发，对 10 个代表性 KOS 管理工具进行调研分析。分析发现，KOS 管理工具正逐渐顺应用户需求及技术发展需求，依托语义网、关联数据、语义集成、语义互操作等技术促进功能创新。未来 KOS 管理工具的建设应注意以下几个方面：灵活设计并配置多类型数据模型、支持可持续发展的管理与维护机制、支持用户分布式协同管理与控制、知识驱动的数据迭代式更新、面向语义的 KOS 重用与进化、建设以 KOS 为核心的知识组织框架等。在功能分析基础之上，笔者从用户使用的角度出发给出了 KOS 管理工具选择建议。

本研究还存在一定的不足，如未将用户的使用体验纳入评估范畴。今后工作中，可总结归纳知识服务过程中来自最终用户、项目管理人员、KOS 数据编辑人员的实际需求，同时借鉴现有 KOS 管理工具的优点，逐步完善自身 KOS 管理工具的建设工作。

参考文献：

[1] 贺德方. 国内外知识组织体系的研究进展及应对策略[J]. 情报学报, 2010, 29(6):963-972.

[2] HEDDEN H. The accidental taxonomist[M]. Medford: Information Today, Inc., 2016.

[3] Knowledge organization systems: an overview [EB/OL]. [2020-09-28]. https://www.clir.org/pubs/reports/pub91/1knowledge.

[4] TUDHOPE D, NIELSEN M L. Introduction to knowledge organization systems and services[J]. New review of hypermedia and multimedia, 2006, 12(1): 3-9.

[5] Taxonomy 101: The basics and getting started with taxonomies [EB/OL]. [2020-09-28]. http://www.kmworld.com/Articles/Editorial/What-Is/Taxonomy-101-The-Basics-and-Getting-Started-with-Taxonomies-98787.aspx.

[6] 白海燕. 传统知识组织工具管理软件的发展与评价[J]. 现代图书情报技术,2008(9):47-52.

[7] Taxonomy software trends. [EB/OL]. [2020-09-28]. https://accidental-taxonomist.blogspot.com/2015/01/taxonomy-software-trends.html.

[8] GANZMARM J. Criteria for the evaluation of thesaurus software [J]. Ko knowledge organiation, 1990, 17(3/4):148-57.

[9] CHAUDHRY A S. Assessment of taxonomy building tools[J]. The electronic library, 2010, 28(6): 769-788.

[10] PEREZ-LEON B, MARTINEZ-GONZALEZ M M. A comparative study of thesauri tools-a perspective from integrability in information systems[C]//FILIPE J, CORDEIRO J. Proceedings of the 6th international conference on Web information systems and technology. Setubal: SCITEPRESS, 2010: 203-206.

[11] MYRSETH P, YANG J J, OVERBY E. Survey on vocabulary and ontology tools: including a methodology for comparing tools: version 1.0[EB/OL]. [2020-09-28]. http://www.semicolon.no/wp-content/uploads/2013/09/Semicolon_Vocabulary-tools-survey_v1.0.pdf.

[12] MARTINEZ-GONZALEZ M M, ALVITE-DIEZ M L. On the evaluation of thesaurus tools compatible with the Semantic Web[J]. Journal of information science, 2014, 40(6):711-722.

[13] MOCHON G, MENDEZ E M, de la FUENTE G B. 27 pawns ready for action: a multi-indicator methodology and evaluation of thesaurus management tools from a LOD perspective[J]. Library hi tech, 2017, 35(1):99-119.

[14] MORSHED A, DUTTA R. Machine learning based vocabulary management tool assessment for the linked open data[J]. International journal of computer applications, 2012 60(9):51-58.

[15] 钱庆,胡铁军,李丹亚,等. 科技知识组织体系素材管理系统的设计与实现[J]. 数字图书馆论坛,2012(12):2-6.

[16] KMWorld 100 companies that matter in knowledge management 2020[EB/OL]. [2020-09-28]. https://www.kmworld.com/Articles/ReadArticle.aspx? ArticleID=135998.

[17] HEDDEN H. Selecting software for taxonomy, thesaurus and ontology management[EB/OL]. [2020-09-28]. http://docplayer.net/162793744-Selecting-software-for-taxonomy-thesaurus-and-ontology-management.html.

[18] Vocbench[EB/OL]. [2020-09-28]. http://vocbench.uniroma2.it/.

[19] PoolParty[EB/OL]. [2020-09-28]. https://www.PoolParty.biz/.

[20] Synaptica KMS. [EB/OL]. [2020-09-28]. https://www.synaptica.com/synaptica-kms/ .

[21] Synaptica Graphite [EB/OL]. [2020-09-28]. https://www.synaptica.com/synaptica-graphite/.

[22] TopBraid EDG vocabulary management [EB/OL]. [2020-09-28]. https://www.topquadrant.com/products/TopBraid-edg-vocabulary-management/.

[23] Mondeca ITM [EB/OL]. [2020-09-28]. https://mondeca.com/.

[24] SKOS Shuttle [EB/OL]. [2020-09-28]. https://skosshuttle.ch/.

[25] TemaTres [EB/OL]. [2020-09-28]. https://sourceforge.net/projects/TemaTres/.

[26] UniLexicon [EB/OL]. [2020-09-28]. https://unilexicon.com/.

[27] MultiTes [EB/OL]. [2020-09-28]. http://www.multites.com/.

[28] 张运良. 大数据服务中知识组织的挑战及应对[J]. 图书情报工作,2020,64(4):88-94.

[29] BLUMAUER A, NAGY H. The knowledge graph cookbook[M]. Vienna: Edition mono/monochrom, 2020.

[30] NISO TR-06-2017, issues in vocabulary management[EB/OL]. [2020-09-28]. https://www.niso.org/publications/tr-06-2017-issues-vocabulary-management.

[31] Shapes constraint language (SHACL) [EB/OL]. [2020-09-28]. https://www.w3.org/TR/shacl/.

[32] 石泽顺,肖明. 基于 PoolParty 的图情学科 SKOS 叙词表构建研究[J]. 图书馆学研究,2017(23):20-30.

[33] SKOS shuttle text extraction management[EB/OL]. [2020-09-28]. https://skosshuttle.ch/tutorials/skosshuttle-textextraction.pdf.

[34] QIN J. Knowledge organization and representation under the AI lens [J]. Journal of data and information science, 2020, 6(1): 3-17.

[35] ISO-THES[EB/OL]. [2020-09-28]. https://lov.linkeddata.es/dataset/lov/vocabs/iso-thes.

[36] 孙海霞, 李军莲, 华薇娜, 等. 科技知识组织体系语义互操作网络协同工作平台设计与实现[J]. 农业图书情报学刊, 2019, 31(1):23-34.

纪姗姗，赵炳容，刘峥. 国外知识组织体系管理工具比较分析与启示[J]. 图书情报工作,2020,64(24):73－83.

作者贡献说明:

纪姗姗:论文撰写,调研设计与发展趋势分析;

赵炳容:参与调研分析;

刘峥:论文选题和写作指导。

Comparative Analysis and Enlightenment of Foreign KOS Management Tools

Ji Shanshan Zhao Bingrong Liu Zheng

National Science Library, Chinese Academy of Sciences, Beijing 100190

Abstract: [**Purpose/significance**] This paper compared and analyzed the current research status and key functions of representative knowledge organization system (KOS) management tools under the development of semantic web, sorts out the development trends of related research, and provided suggestions for the construction of knowledge organization system in various institutions. [**Method/process**] Through online survey, 10 representative foreign KOS management tools were selected, and their development status was compared and analyzed from the aspects of structural definition, project maintenance, data government, interoperability, and operating environment. Based on the survey results, we proposed some suggestions for the construction of KOS management tools in China. [**Result/conclusion**] Relying on the development of technologies such as semantic Web, linked data, semantic integration, and semantic interoperation, the domestic construction of KOS management tools should gradually achieve functional innovation, including flexible design of multiple data models, project management that supports sustainable development, distributed collaborative user management, knowledge driven data updating, semantic-oriented data utilization, KOS as the core of the knowledge organization framework, etc.

Keywords: knowledge organization system (KOS) KOS management tool knowledge management taxonomy ontology